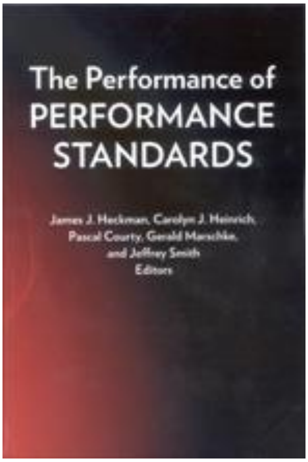
Upjohn Institute Press

A Formal Model of a Performance Incentive System

James J. Heckman
University of Chicago

Carolyn J. Heinrich
University of Wisconsin

Jeffrey Smith
University of Michigan



Chapter 3 (pp. 29-64) in:
The Performance of Performance Standards
James J. Heckman, Carolyn J. Heinrich, Pascal Courty, Gerald Marschke, and Jeffrey Smith, eds.
Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, 2011
DOI: 10.17848/9780880993982.ch3

The Performance of Performance Standards

James J. Heckman
Carolyn J. Heinrich
Pascal Courty
Gerald Marschke
Jeffrey Smith
Editors

2011

W.E. Upjohn Institute for Employment Research
Kalamazoo, Michigan

Library of Congress Cataloging-in-Publication Data

The performance of performance standards / James J. Heckman . . . [et al.], editors.
p. cm.
Includes bibliographical references and index.
ISBN-13: 978-0-88099-292-3 (pbk. : alk. paper)
ISBN-10: 0-88099-292-1 (pbk. : alk. paper)
ISBN-13: 978-0-88099-294-7 (hardcover : alk. paper)
ISBN-10: 0-88099-294-8 (hardcover : alk. paper)
1. Government productivity. 2. Performance standards. 3. Civil service—Personnel management. I. Heckman, James J. (James Joseph)
JF1525.P67P476 2011
352.6'7—dc22
2011007877

W.E. Upjohn Institute for Employment Research
300 S. Westnedge Avenue
Kalamazoo, Michigan 49007-4686

Cover design by Alcorn Publication Design.
Index prepared by Diane Worden.
Printed in the United States of America.
Printed on recycled paper.

3
A Formal Model of a Performance Incentive System

James J. Heckman
Carolyn J. Heinrich
Jeffrey Smith

This chapter presents a model of training center behavior in an environment that includes a generic performance management system for active labor market programs (ALMPs), such as those funded under JTPA and WIA in the United States. The model builds on the work of Heckman, Heinrich, and Smith (2002) and provides more intuition and discussion of the model and its implications, along with some useful extensions.[1] Additionally, our model offers an essential conceptual context for the detailed analyses of the JTPA and WIA programs that follow in the remaining chapters of this volume.

The model we develop assumes that training (or workforce development) centers seek to maximize the present discounted value of earnings (or employment) impacts from the services they provide, as well as, potentially, goals related to the characteristics of participants and to the effort levels exerted by program staff.[2] The JTPA and WIA programs both have formally stated equity (service to particular subgroups) and efficiency (improving labor market outcomes relative to what would have occurred without the program) goals. Our model demonstrates how these objectives interact with a performance standards system based on short-term outcome measures, and we discuss in detail why all of the performance standards systems we know of rely on performance measures based on outcome levels measured in the short term rather than on impacts ("value added") over the long run. The use of short-term outcomes as performance measures has the potential to misdirect activity by focusing training center attention on criteria only loosely (or even perversely) related to long-run net benefits, long-run equity criteria, or both. For example, if program activities encourage further training and schooling, they may reduce employment and earnings in the short run but raise them in the long run.

Most discussions of performance standards focus on cream skimming, sometimes defined as selecting persons into a program who would have done well without it. Anderson et al. (1992) and Barnow (1992) represent early examples of this literature, while Courty, Kim, and Marschke (forthcoming) provide an important recent contribution. In the context of a system of performance standards, cream skimming occurs when training centers serve individuals who will increase their measured performance rather than basing service provision decisions on individuals' expected long-run benefits from participation. In this chapter, we provide a concise formal definition of cream skimming in terms of our model notation, show how performance standards based on participant outcomes in the short run encourage it, and discuss the conditions under which it does or does not lead centers away from the goal of maximizing discounted impacts.

The model we introduce allows responses to performance standards in terms of who gets selected into the program from among the eligible population and how program resources get allocated among participants, as well as strategic responses that seek to increase measured performance conditional on actual performance, such as those considered in Chapter 7. More generally, this model provides a clear conceptual framework within which to think about when (and if) performance standards systems in ALMPs will increase the labor market impacts they produce, and when (and if) they will have unintended consequences due to responses by program staff to the sometimes misguided incentives they provide.

A MODEL OF TRAINING CENTER CHOICES

Training centers face three choices in each period: 1) how many people to train, 2) which particular people to train, and 3) how many resources to devote to each trainee. In the current WIA program, access to "core" workforce development services (such as labor market information and job search assistance) is universal, but access to more intensive levels of service (such as comprehensive assessment and case management, vocational training, and subsidized on-the-job training) still involves some degree of selection by caseworkers in light of the available training resources for individual clients and services (see Social Policy Research Associates [2004] for more institutional details on WIA). Thus, in the WIA context, our model applies to these more intensive services. Adding a requirement that all eligible individuals receive some very small level of service would not change any of the basic results from the model.

For simplicity, we assume that training centers face a completely new cohort of potential trainees in each period; this avoids potential complications associated with training centers making choices about when to serve particular people. More generally, we assume that training centers operate in a "stationary" environment, which means that the center's decision rules do not change over time. Put differently, if a center faces the same set of potential trainees, the same technology for producing trainee outcomes, and the same budget constraint in two different periods, it will make exactly the same choices in those two periods. We assume that the set of potential trainees and the technology stay the same, and, in later sections, that the budget varies only as a result of performance awards.

We ignore the individual application decision in our model and simply assume that the training center can choose to serve any or all of the eligible individuals in each cohort (given its budget constraint). In so doing, we abstract from center choices regarding marketing, outreach, contractor selection, and other factors that might affect who applies to the program, even though the presence or details of a performance standards system might affect these choices.

We also assume that individuals participate or not in the first period of their adult lives, which we denote age zero, and that training takes exactly one period for all trainees. Thus, we implicitly ignore individual choices regarding the timing of training.[3] Each individual has two potential outcomes at each age: a benchmark (or untreated) outcome that arises if the individual does not participate (at age zero), and a treated outcome that arises if the individual does participate (at age zero). In terms of the usual notation, we denote the participant outcomes by $Y_a^1, a = 0, \ldots, A$, where A is the final period of the person's life, and denote the nonparticipant outcomes by $Y_a^0, a = 0, \ldots, A$. The per-period treatment effect equals $Y_a^1 - Y_a^0 = \Delta_a$. The treatment effect can be negative in the short run if, for example, program participation leads to additional schooling or distracts the individual from useful job search.[4] We abstract from potential general equilibrium effects in our discussion.[5] In the language of the treatment effects literature, we make the stable unit treatment value assumption (SUTVA), which means that the treated and untreated outcomes just defined do not depend on who participates or on how many individuals participate.

To allow our analysis to fit into a standard cost-benefit framework, let Y denote earnings; we can easily generalize the model to include employment or education outcomes. The net (of costs) present value of the program impacts (hereinafter just "net impacts") measured at time zero for participant i then equals

$$(3.1) \qquad PV_i = \sum_{a=0}^{A} \frac{\Delta_{ai}}{(1+r)^a} - c_i \, ,$$

where c_i denotes the costs associated with participant i. We assume that (Δ_a, c) varies among individuals but the interest rate r does not.[6]

The model assumes that training centers can apply different amounts of input, e, to each participant. In the context of JTPA or WIA (or similar programs), the input variable represents the value of staff time and the direct costs of the services provided. The inputs affect the outcomes experienced by participants. In particular, input e yields

$$(3.2) \qquad Y_a^1 = f(Y_a^0, e) \, ,$$

at cost $c(e)$, where $c(0) = 0$. The total cost for participant i is given by $c_i = c_i(e_i) + k$, where k denotes a per-participant fixed cost. Note that we allow both the amount of the inputs and the marginal cost of the inputs to vary among participants.

By choosing to model a continuous input e, we abstract from reality on two important dimensions. First, most of the services provided by ALMPs come in discrete chunks of a particular type. For example, JTPA offered, and WIA currently offers, classroom training in occupational skills, job search assistance, and subsidized on-the-job training at private firms, among other services. Classroom training consists of various types of courses, each aimed at a particular occupation and each having a specific duration. The other service types vary in a similar way. Representing this complex mix of discrete services by a continuous input simplifies our model considerably, but at the cost of abstracting from the potential impact of performance standards not only on the amount of services provided but on their type and duration. For example, Heinrich (1999), Barnow and Gubits (2002), and D'Amico and Salzman (2004) (and many others) have argued that a focus on short-term outcomes in performance standards pushes training centers toward providing shorter, less-intensive services such as job search assistance to a larger number of trainees. Second, we ignore the fact that inputs often get allocated to participants dynamically in response to their experiences in particular treatments and in response to their labor market outcomes. For example, at the end of a classroom training course, participants with a job lined up do not receive job search assistance, while those without a job lined up often do. Our simplifying assumption means that our model also fails to capture any effects of performance standards on this dynamic service allocation process.

Given these assumptions, training centers have several degrees of freedom in regard to whom to serve and how many inputs to devote to each participant. First, for a fixed set of inputs, a training center can choose to serve individuals with different (Δ_a, c) combinations. Second, holding the set of participants fixed, the training center can choose the inputs it provides to each participant, which has the effect of changing their potential outcomes given participation (and, necessarily, changing their impacts of participation as well). Third, a training center can choose the number of participants by trading off between the fixed cost k and the variable input cost $c_i(e_i)$.

If the training center maximizes the ex post present value of the net earnings impacts realized by its trainees, it solves a constrained optimization problem. Maximizing the present value of the net impacts corresponds to a social goal of efficiency in the economic sense; it means making the economic "pie" as large as possible while ignoring equity concerns other than those implicit in the program's eligibility rules. In the absence of a budget constraint, the center would simply find the e that maximizes the present value of net impacts for each participant:

$$(3.3) \qquad \hat{e} = \arg\max_{e} \sum_{a=0}^{A} \frac{(Y_a^1 - Y_a^0)}{(1+r)^a} - c(e) - k.$$

In the real world, and in our model, training centers operate under a budget constraint. Let B denote the center budget in each period. With a budget constraint, centers face a trade-off between serving more clients and devoting more inputs to each client. Let $\{1,\ldots,I\}$ be the index set of eligible individuals; put differently, label each eligible person with a number from 1 to *I*. Person *i* has associated variable costs $c_i(e_i)$ and fixed cost *k*. We assume that technology (3.2) is common across persons although this assumption can easily be relaxed.

The training center solves its maximization problem in two steps. In the first step, for each possible set of trainees $S \subseteq \{1,\ldots,I\}$ in the current cohort it determines the optimal choice of inputs to devote to each member of the set. Formally, for each possible set *S*, the center solves the problem

$$(3.4) \quad \max_{e_i, i\in S} \sum_{i\in S}\left[\sum_{a=0}^{A}\frac{\left(Y_{a,i}^{1}-Y_{a,i}^{0}\right)}{\left(1+r\right)^{a}}-c_i\left(e_i\right)-k\right],$$

subject to Equation (3.2), and

$$(3.5) \quad B \geq \sum_{i\in S}\left(c_i(e_i)+k\right).$$

For LaGrange multiplier λ attached to the constraint in (3.5), solving the optimization problem produces the first order condition

$$(3.6) \quad \sum_{a=0}^{A}\left[\frac{\partial f\left(Y_{i,a}^{0},e_i\right)}{\partial e_i}\right]\frac{1}{\left(1+r\right)^{a}}=\lambda\frac{\partial c_i\left(e_i\right)}{\partial e_i}$$

for each observation $i \in S$. Condition (3.6) represents the standard efficiency condition for e_i which sets marginal benefit equal to marginal cost. In the absence of a budget constraint, $\lambda = 1$ at an interior optimum. In general, $\lambda \geq 1$, reflecting the scarcity of the resources available to the center, and the center invests less in each person than it would in the absence of resource constraints.[7]

The second step to solving the overall optimization problem for each cohort consists of comparing the optimal value of the present value of net impacts for each subset *S* and choosing the subset S^* that has the highest one. Formally, write the maximized present value implied by the solution to the constrained optimization problem in Equation (3.4) as $\psi(S,B)$, where this notation shows the dependence of the optimum on both the set S of participants selected and the available budget B. The center chooses the optimal S, which we denote by S^*, so that

$$\psi(S^*,B) = \arg\max_{S} \psi(S,B).$$ [8]

Implementing this ex post optimal solution requires information that both the centers and (to a lesser extent) the individuals themselves do not possess at the time of participation. In particular, they do not know future values of (Y_a^1, Y_a^0), although they may have other information useful for predicting these values. For example, most ALMPs collect information on observable characteristics associated with outcomes from prospective participants, and some may also have access to administrative data on past labor market outcomes.

The available evidence suggests the difficulty of forecasting future Δ_a. In particular, Bell and Orr (2002) show that caseworkers do a very poor job of predicting Δ_a in a program that provides job training to welfare recipients, and Lechner and Smith (2007) show that Swiss caseworkers also do not perform well at this task. Carneiro, Hansen, and Heckman (2003) demonstrate that individuals cannot forecast most of the variation in the earnings impact associated with attending college.

Let J_i denote the center's information set for individual i. Taking into account the lack of complete information, the criterion for *ex ante* optimality for each S becomes

$$\text{(3.7)} \quad \underset{e_i}{\text{Max}} \sum_{i \in S} \left\{ \left[\sum_{a=0}^{A} \frac{E\left(Y_{a,i}^1 - Y_{a,i}^0 \mid J_i\right)}{\left(1+r\right)^a} \right] - c_i(e_i) - k \right\},$$

subject to Equation (3.2), Equation (3.5), and the individual-specific information sets $\{J_i\}_{i \in S}$. For each $S, \{J_i\}_{i \in S}$, B, and r, we may write the solution to this present value maximization problem as $\psi\left(S,B,J\right)$, where $J = \{J_1, \ldots, J_I\}$. The training center seeks to maximize this criterion with respect to S, so that

$$\psi\left(S^*,B,J\right) = \arg\max_{S} \psi\left(S,B,J\right).$$

In this model, training centers adjust at three margins: 1) which eligible individuals become trainees (in WIA, which participants receive intense services), 2) the amount of inputs devoted to each trainee, and 3) the number of trainees. While the exact trade-offs depend on the specifications of the technology for producing outcomes in Equation (3.2), the marginal cost functions, and the level of fixed costs k, a set of intuitive comparative static results follow directly from the model. For example, increasing the slope of the marginal cost function $c(e)$, all else equal, leads centers to increase the number of participants and to serve each one less intensively. Increases in the fixed cost k have the opposite effect, reducing the number of participants and increasing the amount of resources devoted to each one. Individuals with higher marginal costs—i.e., larger values of $c_i'(e_i)$—will, all else equal, receive fewer inputs. Increasing the amount of complementarity between the untreated outcome and the costs in the production function in Equation (3.2) leads centers to devote relatively more inputs to participants with good untreated outcomes. Increases in the budget lead, in general, to both more participants and more inputs per participant. Finally, in a stationary environment, the training center makes the same decision in every period.[9]

MOTIVATING PERFORMANCE STANDARDS

The model in the preceding section assumes that training centers maximize the net present value of impacts and nothing else. In fact, training centers exist in a political context and they employ caseworkers and managers who care about outcomes other than just the maximization of the present value of net impacts. Consider the politicians first. Politicians care about what training centers do. In particular, they care not only about present value maximization but also about other aspects of who, and how many, get served. For example, politicians may care about the absolute number of participants, based on the view that each participant will feel that he or she has received something from the politician, and so, perhaps, will vote accordingly. Politicians may also care about serving members of particular groups whose leaders support them or about serving highly visible individuals, such as those who lose their jobs when a major plant closes; see, e.g., Heinrich (1999) on the role of politics in JTPA contract award decisions.

We can summarize politicians' preferences by the utility function $U_P[\psi(S),N(S),Q(S)]$, where $\psi(S)$ denotes the expected present value of net impacts for participant cohort S (with the other arguments suppressed for simplicity), and $N(S)$ denotes the number of participants ($\leq I$) served in cohort S. $Q(S)$ denotes other qualities of the persons served such as demographic characteristics or their untreated outcomes, with the latter motivated by a desire to serve those least well off in the absence of the program due to equity concerns.

Training center staff members also care about aspects of their work other than just present value maximization. For example, as discussed in Heckman, Smith, and Taber (1996), caseworkers may prefer to serve the most disadvantaged (those with the lowest benchmark outcomes Y_a^0) or members of particular subgroups among the eligible. At the same time, they may prefer to serve fewer individuals than present value maximization would imply if, for example, they get utility from getting to know their clients in depth. Abstracting for simplicity from the fact that training center managers (some of them future politicians) may have different preferences than the line workers they manage, we can summarize the training center utility function by $U_T[\psi(S),N(S),Q(S)]$.

Given either utility function, we can define a constrained optimization problem similar to that defined in the preceding section. The problem consists of maximizing the utility function subject to the technology for producing participant outcomes in Equation (3.2) and the budget constraint in Equation (3.5) through choices about how many people to serve, which ones to serve, and how to allocate the inputs among those served. The equilibrium from this modified optimization problem will differ in simple and intuitive ways from that obtained under present value maximization in the preceding section. For example, if we consider maximizing the training center's utility and if the training center gets disutility from a larger number of participants, then the resulting number of participants will not exceed that chosen under present value maximization. If the training center gets utility from serving some particular group, say unemployed musicians, then it will serve more of them (or at least no fewer of them) under utility maximization than under present value maximization, and so on.

In this setting, performance standards have no role to play, even if the training center has imperfect information about potential outcomes, so long as the utility functions of the politicians and the training center are similar enough (e.g., one is a linear transformation of the other) that they would reach the same solution to the *ex ante* utility maximization problem just described. A role for performance standards emerges when the two utility functions imply different choices. In that event, politicians may want to provide additional incentives to training centers in a way that makes their choices closer to those implied by the politicians' utility function. For example, if the training centers dislike having more participants at the margin and politicians like having more participants at the margin, then politicians will want to introduce performance standards in a way that rewards centers for having more participants.

In one view, this situation, in which the politicians set up the local training centers, which then deliver services within the context of broad rules, represents a classic principal-agent problem. The politicians would like the training centers to choose an optimum that corresponds to their own utility, rather than that of the center staff. See the insightful survey in Dixit (2002) and the references cited therein (as well as the other papers in the related special issue of the *Journal of Human Resources*) for further elaboration of this point.

Politicians can bring center behavior in line with their own preferences in two basic ways.[10] One obvious way consists of specifying the rules governing center behavior so narrowly as to leave center staff with little discretion to do otherwise than as specified by program rules.[11] The fundamental problem with this approach in the context of ALMPs arises from the complexity of the task at hand and the large amount of (often tacit) local knowledge, in the sense of Hayek (1945), required for the task. Most training centers under JTPA and now under WIA serve individuals with quite heterogeneous desires and abilities by matching them with a (wide) variety of services provided either directly by the center or a service provider under contract. Attempting to prespecify the matches between participants and services would require either numbing levels of regulatory detail or a false simplicity that would likely seriously degrade the quality of the resulting matches. Performance standards represent an alternative to micromanagement by the politicians via program regulations. Here politicians define the goals they seek by defining the performance measures and the "reward functions" (described in Chapters 4 and 5 for the JTPA program) that link observed performance to rewards and punishments. In so doing, they allow centers to continue using their local knowledge in choosing whom to serve and how to serve them, while at the same time directing the application of this knowledge toward the politicians' own goals.

An alternative view from a principal-agent perspective assumes that politicians and training centers share the same preferences about whom to serve and how to serve them, but differ in their desired level of training center effort. Put simply, politicians would like training center staff to work very hard, while training center staff would like to consume some on-the-job leisure. This view links to the literature on piece rates, performance-based compensation for CEOs, and other types of incentives often found in private firms; this literature focuses almost exclusively on methods for tying remuneration to measured output as a means to increase unobserved effort (see Prendergast [1999] for an able survey, Lazear [1995] for a book-length treatment, and Chapter 5 for additional discussion in our context). We could incorporate this view into our model by breaking inputs e into two components, one representing staff effort and the other representing other inputs. The sum of staff effort would then enter negatively into the training center utility function, capturing the negative direct effect of staff effort; at the same time, staff effort would have an indirect positive effect on center utility through its effect on the present value of net impacts.

A third view of performance standards emphasizes the information they provide rather than their role in solving (or attempting to solve) one or the other of the principal-agent problems just described. In this view—discussed, for example, in Smith (2004)—performance standards represent quick and dirty impact evaluations. They provide valuable feedback to training center managers and staff (and to politicians) about their progress at meeting equity and efficiency goals—feedback that, as Lechner and Smith (2007) note, training centers often otherwise never receive, because they rarely interact with those they serve after they serve them. Because the feedback comes quickly, it allows rapid responses to changes in performance due to changes in program operation, local economic conditions, or other factors. Our model does not capture this role for performance standards in a formal way; doing so would substantially complicate the model, as it would need to incorporate training center learning in response to the information provided by the performance measures. We note in Chapter 9 some implications of thinking about performance measures in this way for the interpretation of the evidence on the correlation between performance measures and program impacts. Rather obviously, if performance measures exist to proxy long-run impacts, their correlation with those impacts becomes the paramount measure of their value.

All three views about the role of performance standards appear in the literature; in our view, all three have empirical relevance. In this chapter, we focus our model mainly on the first of the three roles; generalizing our analysis to include the second is straightforward.

PERFORMANCE MEASURES IN PRACTICE

In practice, most performance measures M consist of participant outcomes in the short run. The focus on the short run stems from the desire to provide prompt feedback to program managers and politicians. Feedback that arrives years after the corresponding actions by program staff does little to either motivate or inform. The focus on observed outcomes (i.e., did the participant get a job soon after finishing the program?) rather than estimated impacts (i.e., did the participant do well in the labor market relative to what would have transpired had he or she not participated?) has several motivations. First, evaluations (experimental or nonexperimental) that seek to estimate impacts by estimating the counterfactual outcomes of participants take a long time, typically on the order of years, to carry out. Even short-run impact estimates require considerable time to collect and prepare the necessary outcome data on participants (and, if required, a comparison group) for use in econometric analyses. Second, performance measures based on impacts often generate controversy, either because of uncertainty about the econometric method utilized, in the case of nonexperimental methods, or politically, in the case of random assignment. Finally, performance measures based on outcome levels generally cost much less to produce than measures based on impacts. As the literature on evaluating active labor market programs makes clear (see, e.g., Heckman, LaLonde, and Smith 1999), impact evaluation poses demanding technical problems that typically require the assistance of expensive experts and make the process difficult to automate. In contrast, measuring outcomes presents much simpler problems and, once established, becomes a routine data collection and processing exercise. Costs matter, because an expensive performance management system, even if it accomplishes something, may not accomplish enough to justify the expense.

The most common performance measure consists of employment rates in the period immediately following program participation. In terms of our notation, we can represent this measure as

$$(3.8) \quad M(S_0) = \frac{1}{N(S_0)} \sum_{i \in S_0} 1(Y_{1,i}^1 > 0) = \frac{1}{N(S_0)} \sum_{i \in S_0} E_{1,i}^1 ,$$

where the first subscript on $Y_{1,i}^1$ denotes age 1, the 0 subscript on S_0 indicates the current cohort of trainees; $1(\cdot)$ denotes the indicator function, which takes a value of 1 when its argument holds and a value of 0 when it does not; and $E_{1,i}^1 = 1(Y_{1,i}^1 > 0)$ is a dummy variable for employment (defined here as positive earnings) at age 1 (the first period after participation). A slightly different formulation captures measures such as those in JTPA and WIA that consider wages or earnings conditional on employment. For example, in our notation, a measure based on earnings conditional on employment becomes

$$M(S_0) = \frac{1}{\sum_{i \in S_0} 1(Y_{1,i}^1 > 0)} \sum_{i \in S_0} 1(Y_{1,i}^1 > 0) Y_{1,i}^1 = \frac{1}{\sum_{i \in S_0} 1(Y_{1,i}^1 > 0)} \sum_{i \in S_0} E_{1,i}^1 Y_{1,i}^1 .$$

In this performance measure, the indicator function serves to condition the average on the set of participants with positive wages or earnings.

As described in Chapters 2 and 4, not all performance measures focus on labor market outcomes. Some capture the receipt of various educational credentials, such as the GED, or, in the WIA program, customer satisfaction measures based on surveys (where customers include both participants and the firms that might hire them). Smith, Whalley, and Wilcox (2010a,b) discuss and provide evidence on the value of such participant self-evaluation measures. For many years, the WIA performance standards system also included a measure based on before-after earnings changes, which took advantage of the well-known preprogram dip in mean earnings among participants to produce the (highly misleading) appearance of large program impacts. Heckman and Smith (1999) document the dip using the National JTPA Study data and show its implications for, among others, the before-after estimator of program impacts. Incorporating any or all of these alternative performance measures into our model requires no conceptual innovations.

REWARD FUNCTIONS IN THE MODEL

The reward function *R*(*M*) links observed performance on the performance measures to rewards and (much more rarely) punishments for training centers. In most performance systems for ALMPs around the world, rewards come informally, through praise and recognition. In contrast, Chapter 4 documents that rewards in the JTPA performance standards system took the form of budgetary increments determined by formal rules. The WIA system lies somewhere in between, with monetary bonuses awarded to states that were probably more consequential for the recognition they accord than the funds they provide (see, for example, Heinrich [2007]).

The simplest system assigns a discrete reward *R* to centers meeting a defined standard in terms of observed performance. In terms of our notation, this implies the reward function

$$(3.9) \qquad R(M(S)) = \begin{cases} R \text{ if } M(S) > \tau; \\ 0 \text{ otherwise.} \end{cases}$$

Here τ denotes some fixed level of attainment on the performance measure, such as 60 percent of former participants employed in the period after training. Training centers that exceed that level get the reward and those that fall below it do not. As noted in Chapter 4, many more complicated reward functions exist (or have existed) in practice, including functions that reward relative rather than absolute performance, functions that require certain levels of performance on multiple measures, and functions that reward marginal improvements in performance above τ. Extending our model to incorporate such reward functions is straightforward.

For nonbudgetary rewards, it makes sense to put the reward function directly in the training center utility function, as in

$U_T[\psi(S),N(S),Q(S),R(M)]$ where we leave the dependence of M on training center choices regarding whom to serve (S) and how to allocate the inputs e among those they serve implicit. Almost by definition, we assume a positive partial derivative.

In the case of budgetary rewards, things become a bit more complicated. In that case, as in the JTPA and WIA programs, recognition remains part of the reward for good performance, and thus it makes sense to keep the reward function in the utility function. At the same time, receiving a budgetary reward changes the underlying choice problem in the next period by allowing the center to serve more individuals, to direct more inputs to those it would have served in any case, or both. Exactly how the additional budget affects choices depends in part on whether individuals served with the reward money count toward measured performance in the following period. If they do not, this allows centers to focus on satisfying their preferences regarding $Q(S)$ with the reward money; for example, they might devote additional attention to the "hard to serve" among their eligible population. In the JTPA and WIA systems, individuals served with reward money do not, in fact, count toward measured performance; nonetheless, later in this chapter we assume for simplicity in our discussions of dynamics that they do, so that we can simply add the reward money R to the original budget B and proceed as before.

THE EFFECT OF PERFORMANCE STANDARDS ON CENTER BEHAVIOR: CREAM SKIMMING

This section outlines the implications of our model for center behavior in the presence of a performance standards system that uses the mean earnings (including the zeros) of participants at age "1" (i.e., in the period after program participation) as the performance measure and includes a reward function that increases in measured performance. To make things even simpler, in this section we suppose that training center utility depends only on the present value of net impacts and the reward, corresponding to the utility function $U_T[\psi(S),R(M)]$.

Adding the basic performance management system just described to the model changes all of the first order conditions. In each case, when choosing how many to serve, whom to serve, and how to allocate inputs, the training center now considers the effect of the choice on both $\psi(S)$ and $R(M)$. Thus, when it evaluates a potential participant, instead of just considering their expected discounted stream of net impacts, the center also considers their expected earnings in the period immediately following participation. A potential participant who wants to enroll in, say, a two-year vocational training program after getting a GED with the help of the program becomes less attractive in the presence of the performance standards system laid out in the preceding paragraph. On the other hand, a potential participant who will find a job with high earnings one or two weeks earlier with the program's help but otherwise derives no benefit from it becomes more attractive. This new emphasis on short-term outcomes in choosing whom to serve and how to allocate inputs leads to the common criticism that performance standards systems generate cream skimming (see, e.g., Barnow and Smith [2004]; Radin [2006]; Trutko et al. [2005]; GAO [2002]).

Before turning to a discussion of the effects of cream skimming on the efficiency and equity of training center operation, we now formally define cream skimming. The policy literature often defines cream skimming rather casually to mean serving the least "hard to serve" among the eligible population. In terms of our model's notation, the simple performance system described above creates an incentive to serve persons with high values of $Y_{1,i}^{1}$, regardless of whether that high value results from a high value of $Y_{1,i}^{0}$ or a high value of $\Delta_{1,i}$.

The existing literature is vague about whether cream skimming should be defined in terms of $Y_{1,i}^{1}$ or $Y_{1,i}^{0}$. Our model, and the logic of performance measurement systems more generally, suggests a definition in terms of $Y_{1,i}^{1}$, as it is $Y_{1,i}^{1}$ that the performance standards incentivize training centers to take account of in their decisions.[12] Of course, given the evidence in Bell and Orr (2002) that caseworkers do reasonably well at forecasting $Y_{0,i}^{1}$ and reasonably poorly at forecasting $\Delta_{1,i}$, as well as the evidence in Heckman, Smith, and Clements (1997) that, at least for the JTPA program, most of the variance in $Y_{1,i}^{1}$ corresponds to variance in $Y_{1,i}^{0}$ (or, put differently, the variance of $\Delta_{1,i}$ is small relative to that of $Y_{1,i}^{0}$), the distinction may not matter much empirically.[13]

We now consider the implications of cream skimming for program equity and efficiency in the context of our model. To keep the discussion simple, for the purposes of this section we make the simplifying assumption that untreated outcomes do not vary with age, so that $Y_{0,i}^{a} = Y_{0,i}$ for all "a." This assumption allows us to summarize equity concerns in a single number, where we define equity as serving those with the lowest values of $Y_{0,i}$ among the eligible.

Consider first an important special case. As noted in Heckman (1992); Heckman, Smith, and Clements (1997); and Djebbari and Smith (2008), conventional models of program evaluation assume that $Y_{a,i}^{1}$ and $Y_{a,i}^{0}$ differ by a constant, so that $\Delta_{a,i} = Y_{a,i}^{1} - Y_{a,i}^{0} = \Delta_{a}$ for all i. Put differently, they assume that everyone has the same impact of treatment—the so-called common effect model.[14] In the common effect world, a high $Y_{1,i}^{1}$ goes hand in glove with a high $Y_{1,i}^{0}$, and picking persons with a high $Y_{1,i}^{0}$ helps toward satisfying Equation (3.8). In the context of our model, the common effect assumption simplifies the production function in Equation (3.2) to

$$(3.2') \qquad Y_{a}^{1} = f(Y_{a}^{0}, e) = Y_{a}^{0} + \Delta_{a},$$

which removes the input choice decision from the problem (and with it the variable cost portion of the cost function).

Assuming equal fixed costs for all trainees, training centers in a common effect world serve only the individuals at the top of the distribution of untreated outcomes among the eligible. In this world, the discounted stream of impacts does not depend on who gets served, leaving the center free to maximize their measured performance. Thus, in the common effect world, our simple performance standards system has no effect on economic efficiency but very unattractive equity effects.

A mild generalization of the common effect world illustrates another important point. Suppose that all individuals share the same discounted sum of impacts

$$\Delta_{d,i} = \left[\sum_{a=0}^{A} \frac{\left(Y_{a,i}^{1} - Y_{a,i}^{0}\right)}{\left(1+r\right)^{a}} \right],$$

but that the timing of the impacts varies. To keep things very simple, suppose that one random half of the eligibles have impacts of zero in periods 0 and 1, the period of participation and the period just after, while the other random half has impacts of zero in period 0 but positive impacts in period 1 and all future periods, with the impacts just a bit smaller than those in the first group so that the discounted sum comes out the same. Training centers will now prefer, at the margin and to the extent that they can identify them, the individuals with the earlier impacts. Put differently, given the same value of $Y^0_{1,i}$, an individual with a positive impact the period following participation adds more to a center's measured performance than an individual with no positive impacts until the second period after participation. Thus, performance measures based on outcomes in period 1 encourage the provision of services that yield quick improvements in outcomes relative to later improvements in outcomes, even conditional on the same discounted sum. At the margin, the center would even trade off some of the discounted impacts to get a larger impact in the period of performance measurement, an incentive that those who complain about an overemphasis on low-intensity "work first" strategies clearly have in mind.

Another simple model inspired by Heckman, Smith, and Clements (1997) assumes independence between impacts and untreated outcomes, while continuing to assume no variable inputs and constant fixed costs. The production function becomes

$$(3.2') \qquad Y^1_{a,i} = f(Y^0_{a,i}, e) = Y^0_{a,i} + \Delta_{a,i},$$

with $\Delta_{a,i}$ independent of $Y^0_{a,i}$ for $a = 1, \ldots, A$. In this world, in the absence of performance standards the training centers rank individuals by their discounted impacts and start serving individuals from the top, continuing down the distribution until the budget runs out from paying fixed costs. In contrast, adding in our simple performance standards system makes the problem two-dimensional, with centers now serving those individuals with good present values of impacts and good outcomes in the period following participation, whether due to a large impact or to a good untreated outcome (or both) in that period. As in the common effect model, the introduction of performance standards leads centers to move toward serving individuals who, on average, have better outcomes in the untreated state. Thus, once again, it has problematic equity effects. In this model, the performance standards clearly reduce efficiency, as centers will now, at the margin, implicitly trade off discounted impacts for good untreated outcomes in the period of performance measurement.

Finally, consider the same simple world as in the previous example, with production function (3.2') and so on, but assume that impacts and untreated outcomes have rank correlation γ. We have already considered the case of independence, which implies a rank correlation of zero. With a positive rank correlation equal to one, training center behavior does not change at all with the introduction of performance standards because, given our simplifying assumptions, the same individuals have both the largest impacts and the largest outcomes in the period following participation. More generally, with a positive rank correlation of less than one, we expect relatively small reductions in both the equity and efficiency associated with training center choices. This is because some individuals with good discounted impacts but bad outcomes in period 1 get dropped in favor of individuals with good untreated outcomes in period 1 but smaller discounted impacts; at the same time, most individuals either participate or not both with and without the performance standards. In contrast, with a negative rank correlation, the introduction of performance standards should lead to greater losses on both dimensions, as there is more scope for training centers to trade off impacts and outcomes. A negative correlation implies more individuals below the cutoff (in terms of discounted impacts in a world without performance standards) with high values of the untreated outcome, who can therefore add substantially to measured performance in the world with the performance standards.

If we now undo some of our simplifying assumptions by restoring variation in untreated outcomes over time and in marginal inputs costs, the model becomes much more complicated but the same intuition applies. In general, if we start from a situation where training centers care only about discounted net impacts, adding performance standards to the model reduces efficiency. The common effect case constitutes an interesting but empirically irrelevant exception. The effects on equity depend on the correlation between impacts and untreated outcomes in the period of performance measurement. With no correlation or a positive correlation, performance standards lead to cream skimming, implying negative equity effects.

The final relaxation of our assumptions occurs when we return to assuming that the training center utility function includes not just discounted impacts and performance rewards but also trainee characteristics and the number of trainees. With this utility function as the starting point, it becomes possible to describe cases in which performance standards increase efficiency. For example, if frontline workers have the "social worker mentality" described in Heckman, Smith, and Taber (1996) and prefer to serve those with the lowest untreated outcomes, and if impacts have a positive correlation with untreated outcomes as suggested in Heckman, Smith, and Clements (1997), then performance standards based on short-term outcomes may increase efficiency compared to the status quo, even though they reduce efficiency relative to the case of net impact maximization.

In the end, the effect of introducing performance standards on efficiency becomes an empirical question, as it depends on the relationship between impacts and untreated outcomes, on the relationship between short-run and long-run impacts, and on the extent to which training centers pursue objectives other than maximizing discounted impacts in a world without performance standards. Chapter 6 considers how to measure cream skimming and discusses the available empirical evidence. Chapter 9, as well as Heckman, Smith, and Clements (1997), considers the relationship between outcomes and impacts. Hotz, Imbens, and Klerman (2006); Lechner, Miquel, and Wunsch (2004); and Dyke et al. (2006) (among others) consider the relationship between short-run and long-run impacts.

STRATEGIC RESPONSES TO PERFORMANCE STANDARDS

Until now, we have not considered strategic responses of the sort documented in Chapters 7 and 8 in our model. This section considers "static" strategic responses related to measurement; we consider dynamic responses related to the manipulation of termination dates later on in the chapter after introducing the dynamic version of our model. The static responses we have in mind include the strategic enrollment decisions in the JTPA program documented in Doolittle and Traeger (1990) and the apparent manipulation of the telephone surveys originally used to measure employment shortly after termination in the JTPA program, which the USDOL sought to end with stricter procedural directives.[15] In regard to the former, only individuals formally enrolled in JTPA (or, in WIA, those receiving more than core services), count for the performance measures. The evidence makes it clear that, in the early years of JTPA prior to the tightening of the rules on enrollment, training centers sometimes provided services to individuals but did not formally enroll them until their prospects for contributing positively to the center's measured performance looked good. In regard to the latter, until response rate requirements appeared (and then, later, administrative data replaced telephone surveys for this outcome), centers appeared to selectively survey their recent terminees with the goal of maximizing their measured performance.

A simple extension of the model to incorporate this strategic behavior begins by making a distinction between actual performance and measured performance. Actual performance, denoted by $M(S)$, consists of what the performance measure would equal if measured by a neutral outsider objectively applying program rules, while measured performance, denoted by $M^*(S)$, consists of performance as measured and reported by the training center. Now assume that centers can incur some cost to improve their measured performance. Formally, let

$$M^*(S) = g[M(S),c_s],$$

where c_s denotes the cost of manipulating the measured performance numbers in terms of both staff time and effort and the present value of any political fallout from doing so. Further assume positive first derivatives, as well as a negative second derivative with respect to c_s, so that additional costs increase measured performance but at a decreasing rate. This production function will likely differ among training centers depending on the types of services they provide and the honesty of their staff (which affects their psychic costs of strategic behavior and thus the compensation they must receive for engaging in it). For example, under JTPA, centers that specialized in job search assistance and subsidized on-the-job training at private firms, both of which provide clear signals of likely success at obtaining employment, may have had an easier time manipulating enrollment decisions than centers that provided more classroom training, where employment outcomes typically do not get realized until course completion but where payments to providers may have necessitated enrollment.

Incorporating these costs into the model by replacing actual performance in the reward function with measured performance as determined by this production function adds an additional first order condition to the solution of the training center optimization problem. Training centers now select c_s to balance the marginal benefits and marginal costs of altering their measured performance relative to their actual performance. Because their actual performance enters the production function for measured performance and the cross-partial of the production function need not equal zero, and because centers will choose to spend real resources on manipulating their measured performance, the possibility of strategic misrepresentation also alters centers' actual performance.

PERFORMANCE STANDARDS AND BUDGETARY DYNAMICS IN A TWO-PERIOD MODEL

We now consider the dynamics that arise in our model when training centers receive budgetary rewards for performance (as with the bonus awards under WIA). In particular, assume that reward R augments the center budget for the next cohort of trainees but cannot be used as direct bonuses to center administrators or line workers. The possibility of receiving a budgetary reward directs attention toward the short-run goal of maximizing performance on $M(S_0)$, and may or may not serve to maximize the present value of net impacts $\psi(S,B,J)$ for the current cohort of participants. We begin with an analysis of a model for a training center that serves only two cohorts of trainees, with the first served in period 0 and the second served in period 1. This simple model provides a useful point of departure for the more complicated model we analyze in the next section.

In this context, the incentives provided by performance standards create a new intertemporal dynamic. Decisions by the center today affect the quality and quantity of participants in the first period as well as the resources available to the center to serve the second period cohort. In addition to this intertemporal connection, the center's decision problem changes in character because center performance $M(S_0)$ is a random variable as of date 0. Thus, when making decisions in period 1, the center faces a fixed budget B for period 1 but a stochastic budget for period 2, call it $\tilde{B}$, with the budgetary uncertainty resolved only after the center chooses S_0 and e in the first period and the resulting labor market outcomes for the first period participants are revealed.

Assuming a simple reward function that pays out B for performance above a threshold earnings target, we can write

$$\tilde{B} = \begin{cases} B \text{ if } M(S_0) < \tau; \\ B + R \text{ if } M(S_0) \geq \tau. \end{cases}$$

In this simplified two-cohort model, the center picks S_0 to maximize

$$(3.10) \quad U\left[\psi(S_0, B, J), N(S_0), Q(S_0)\right] + \frac{1}{1+\rho}\Pr\left[M(S_0) \geq \tau \mid S_0\right] \max_{S_1^1} U\left[\psi(S_1^1, B+R, J), N(S_1^1), Q(S_1^1)\right] + \frac{1}{1+\rho}\Pr\left[M(S_0) < \tau \mid S_0\right] \max_{S_1^0} U\left[\psi(S_1^0, B, J), N(S_1^0), Q(S_1^0)\right],$$

where $1/(1+\rho)$ is a discount rate, S_1^1 denotes the participants selected in the second period if $M(S_0) \geq \tau$, and S_1^0 denotes the participants selected in the second period if $M(S_0) < \tau$.

Centers can solve this maximization problem in two stages. For the second period cohort, there are two possible states, corresponding to whether the first cohort succeeds or fails relative to the performance standards. In the first stage, the center solves the second period optimization problem for both possible budgets. Given these optimal values, in the second stage it picks S_0 and e_0 to maximize the criterion Equation (3.10) given the values for the second period selected in the first stage. Heuristically, if S_0 were a continuous variable, and Equation (3.10) were differentiable in S_0, the first order condition for S_0 would be

$$0 = \frac{\partial U\left[\psi(S_0, B, J), N(S_0), Q(S_0), R(S_0)\right]}{\partial S_0} + \frac{\partial \Pr\left[M(S_0) \geq \tau \mid S_0\right]}{\partial S_0} \left\{ \max_{S_1^1} U\left[\psi(S_1^1, B+R, J), N(S_1^1), Q(S_1^1), R(M\left\{S_1^1\right\})\right] - \max_{S_1^0} U\left[\psi(S_1^0, B, J), N(S_1^0), Q(S_1^0), R(M\left\{S_1^0\right\})\right]\right\}.$$

The first term reflects the value of S_0 in raising the current utility of the training center. The second term captures the dynamic effect of budgetary rewards, which link current period performance to future period utility. This effect equals the marginal effect of S_0 on the probability of winning the award times the increase in center utility from winning.[16] Of course, the actual solution is more complicated because the criterion is not differentiable in S_0. A similar condition holds for the choice of inputs e_0 in the first period. As the two-cohort model includes, by definition, no third cohort, the center's second period choices have no intertemporal aspect.

In this two-cohort model, performance standards may distort center choices in two ways. The first consists of the static distortions already discussed. The second results from the intertemporal linkage induced by the budgetary rewards. For certain values of the parameters, the center may choose to trade off first period utility for second period utility by distorting its first period choices so as to obtain the performance reward and thereby a larger budget in the second period. This scenario becomes more likely as the reward increases and $\rho > 0$ decreases, because both of these increase the value in the first period of having a larger budget to spend in the second period. The substantive importance of this scenario also depends on having a sufficiently small positive (or even negative) correlation between the performance measure and discounted impacts. As discussed above, if discounted impacts have a large positive correlation with the performance measure, the center does not face much of a trade-off between the two.

PERFORMANCE STANDARDS AND BUDGETARY DYNAMICS IN A STATIONARY ENVIRONMENT

This simple two-cohort model abstracts from the fact that training centers serve multiple cohorts over many time periods. In this section, we examine our model under the extreme (but analytically simple) assumption that training centers operate forever in a stationary environment; that is, other than the potentially time-varying budget induced by the performance system, the key aspects of the center's decision problem, such as the distribution of eligibles, the budget constraint in

Equation (3.5), and the production function for treated outcomes in Equation (3.2), all remain the same.

In each period, training centers can be in one of two states: 1) in receipt of a bonus *R*, so that they have budget *B*+*R* to spend in the current period, or 2) without the bonus, so that they have budget *B*. The budget in each period depends on the center's choices regarding *S* and *e* in the preceding period. As the model assumes a stationary environment and only two states of the world, the centers face what the technical literature calls a Markovian decision problem. In this type of decision problem, the center's optimal decision depends only on which of the two budget states it occupies in the current period; as a result, the choices *S* and *e* require subscripts for the state of the world but not for the time period.

Let V_0 denote the center's value function with budget *B* in the current period, and let V_1 denote the value function for a center budget *B* + *R* in the current period. Then,

$$V_0 = \arg\max_S U\left[\psi(S,B,J), N(S), Q(S)\right] + \frac{1}{1+\rho}\Pr\left[M(S) \geq \tau\right]V_1$$
$$+\frac{1}{1+\rho}\Pr\left[M(S) < \tau\right]V_0 .$$

Similarly, we have

$$V_1 = \arg\max_S U\left[\psi(S,B+R,J), N(S), Q(S)\right]$$
$$+\frac{1}{1+\rho}\Pr\left[M(S) \geq \tau\right]V_1 + \frac{1}{1+\rho}\Pr\left[M(S) < \tau\right]V_0 .$$

It follows from the usual assumptions about the utility function that $V_1 > V_0$; in other words, centers prefer to have more resources available. The optimal choice of *S* depends on the rewards, the preferences, and the constraints facing centers.

We now present some intuitive analysis of some of the effects of the incentives induced by the performance standards in our simple dynamic model. First, let P^{01} indicate the probability of not receiving a reward in one period and receiving one in the next period and let P^{11} indicate the probability of receiving a reward in two consecutive periods. As having more resources makes it easier to attain all center objectives, including meeting the performance standards, it follows that $P^{11} > P^{01}$. Performance standards with budgetary rewards impart a value to incumbency.

Second, the analysis of the two-period model carries over in part in this more general setting. With sufficiently large R, sufficiently small ρ, and sufficiently misdirected performance incentives (incentives not aligned with maximizing the present value of net impacts), centers may sacrifice utility in the current period in order to obtain the budgetary reward and thereby increase their utility in the next period.

Third, consider the special case of centers that, in the absence of the performance standards, care only about maximizing the present value of net impacts (and thus not about $N(S)$ or $Q(S)$). For certain values of the parameters, such centers may divert resources away from that goal in low budget (nonreward) periods. They will do so in order to get the budgetary reward in the following period, which can then be spent on generating a larger total discounted stream of net impacts than would period-by-period net impact maximization.

Fourth, continuing with the same special case, with a sufficiently low probability of attaining the reward threshold and a sufficiently high reward R, the introduction of performance standards can lower the aggregate net impacts produced by all centers. Unsuccessful centers divert their activities away from productive uses and toward meeting the performance targets but reap no budgetary rewards. Successful centers produce more net impacts in the next period because they have more resources, but these additional impacts may not suffice to make up for the reductions in impacts in the current period from the centers that failed to reach the performance targets despite diverting resources away from present value maximization. If the increases in the present value of net impacts from the successful centers do not exceed the decreases from the unsuccessful centers, then aggregate output falls.

PERFORMANCE STANDARDS AND BUDGETARY DYNAMICS: THE REAL WORLD

As described in Chapters 2 and 4 and in more detail for WIA in Heinrich (2007), both JTPA and WIA built budgetary rewards into their performance systems. However, in both programs, the budgetary awards did not simply get added to the budget for the following period, as we have assumed for simplicity in our model. Instead, in JTPA, individuals served using the budgetary reward money did not count against the performance measures in the following period. This allowed centers to focus these funds on, in most cases, the hard to serve. In the context of our model, they spent them to maximize their utility in terms of $\psi(S,B,J)$, $Q(S)$, and $N(S)$ without worrying about $R[M(S)]$. Similarly, under WIA, most states spend their bonus funds on program improvements, either in the form of new initiatives or enhancements to current program infrastructure and services (e.g., trying to develop learning programs that engage dropouts, at-risk youth, and disadvantaged adults, improving outreach activities, and so on). Taking into account the realities of the programs does not change the basic dynamics outlined in the preceding two sections. For certain parameter values, centers might still find it optimal to take a hit in terms of current period utility in order to obtain the reward money, provided the additional utility obtained in the next period more than makes up for the loss when discounted back to the current period.

Moreover, with economies of scale in center operations, the real world budgetary rewards in JTPA and WIA may still generate the sort of incumbency effects described in the preceding section. Centers in JTPA and WIA could spend their reward money in part on administrative costs. If, for example, they would buy productivity-increasing office equipment or hire better managers with this money, this should spill over to the participants served under the regular budget B. In the model, we could capture this by making $c(e)$ declining in $N(S)$ to reflect economies of scale that arise from using physical or human capital more intensively.

STRATEGIC BEHAVIOR IN A DYNAMIC ENVIRONMENT

Another dynamic incentive induced by performance standards arises when centers have some flexibility over the period in which their participants count for the purposes of the performance measures. As documented for JTPA in Chapter 7, centers have some flexibility in the timing of termination decisions for their participants. Thus, around the end of one period (a program year in JTPA or WIA) they have some choice regarding whether to count particular participants in the current period or in the next period. When centers face nonlinear reward functions, like the simple one considered above in which the center receives R for attaining some threshold level of performance τ but has no incentive to do better at the margin either above or below τ, centers above the cutoff in the current period will want to move "good" participants to the following period. In contrast, a center with performance well below τ in the current period will want to move "bad" participants who might otherwise finish participating in the next period back to the current period, where they can do no harm to a "lost cause."

Our formal model does not capture this particular strategic response to performance standards, though we could modify it to do so. The simplest change would add the capacity for centers to count some current period participant realizations against performance in the following period. Doing so would have some per-participant cost that increased with the number of current period participants counted in the next period. The increasing per-participant cost captures the effect of rules, such as those in JTPA and WIA, that govern how long centers can keep participants "on the books" without spending money on them, as well as the fact that some program services (such as occupational training courses) have fixed durations, thus limiting flexibility. A set of more complicated changes to the model would also allow centers to bring forward good outcomes from the following period.

SUMMARY AND CONCLUSIONS

This chapter has laid out a (relatively) simple model of training center behavior in the context of a generic performance standards system similar to those used in active labor market programs around the world. Such performance standards systems have a variety of justifications, including aligning center behavior with the preferences of politicians who fund but do not operate them, solving a principal-agent problem by increasing the effort levels of center staff, and providing "quick and dirty" pseudo-evaluations of the extent to which the program improves the labor market outcomes of participants.

This model clarifies the ways in which performance standards systems affect center behavior, and the conditions under which those changes will increase the earnings or employment gains from participation, as well as the conditions under which they have positive or negative equity effects by changing the set of persons served among the eligible. We clarify the discussion of cream skimming in the literature in the context of our model. We show that the effects of introducing a performance standards system depend in part on center preferences and in part on empirical parameters such as the correlation between the short-term participant outcomes typically utilized as performance measures and the long-term behavioral impacts that represent the real goal of most programs.

We extend our simple model to include strategic behavior by training centers seeking to "game" the performance measures, whether by playing tricks with measurement in a static context or by manipulating the period in which participants count against the performance measures in a dynamic context. We examine two dynamic versions of our model in which centers receive budgetary rewards in each period based on their measured performance in the prior period. In this dynamic context, further distortions of center choices can result, as they may have an incentive to trade off utility (and perhaps discounted net impacts) in one period to achieve high performance and thus a larger budget in the following period.

Overall, this model provides a framework for thinking about the effects of performance standards on organizational behavior in the context of ALMPs and in broader contexts such as schools. In so doing, it motivates and provides a theoretical context for the empirical investigations presented in the later chapters of this book. Finally, in our view, the analysis in this chapter, along with the empirical evidence presented elsewhere in the book, suggests that policymakers who have mandated such systems, as well as the administrators who have determined their details and undertaken their practical implementation, have often failed to appreciate the difficulty of designing a performance system that makes things better (rather than making them worse), as well as the dependence of the effects of performance standards systems on empirical parameters that remain generally unknown and little investigated.

Notes

1. See Hanushek (2002) for a discussion of accountability systems in education based on performance standards at the teacher and school level. See, for example, Barnow (1992) and Barnow and Smith (2004) for additional discussions of performance standards in publicly provided training programs.
2. Wilson (1989) and Dixit (2002) discuss conflicts regarding the objectives of programs as outcomes of a political process.
3. On the issue of the timing of participation, see, for example, Sianesi (2004), Heckman and Navarro (2007), and Fredrikkson and Johansson (2008) and the papers cited therein.
4. It can also be negative in the long run, as indeed it was for male youth in JTPA. Bloom et al. (1993), Bloom et al. (1997), and Orr et al. (1996) provide more detail regarding the experimental impact estimates from the National JTPA Study.
5. Heckman, Lochner, and Taber (1998) present evidence on the importance of general equilibrium effects in evaluating large scale educational programs and Lise, Seitz, and Smith (2004) provide evidence of such effects in an earnings supplement program.
6. Note that *r* may represent a social discount factor.
7. We assume interior solutions. Sufficient conditions for an interior solution are concavity of Equation (3.2) in *e* for all $Y^0_{a,i}$, convexity of $c_i(e_i)$ for each *i*, and Inada conditions on both cost and technology. For some *S*, the constraint in Equation (3.4) may be slack (that is, $\lambda = 1$ can be obtained).
8. There may be more than one *S* that qualifies. If so, we assume the training center picks one of them at random.
9. There is an additional stage to the allocation process that we do not consider, namely, the allocation of the overall program budget among centers. From the standpoint of economic efficiency, the budget should be allocated to equate returns at the margin for all centers.
10. We ignore other methods for aligning training center behavior with politicians' preferences, such as developing a professional culture among caseworkers. See, for example, Wilson (1989) for further discussion.
11. See the related discussion in Sosin (1986) regarding the interaction between rules and caseworker discretion in the old Aid to Families with Dependent Children (AFDC) program.
12. In thinking about cream skimming from a policy perspective, two other facts should be kept in mind. First, as shown in Chapter 6, even if cream skimming occurs, the operation of program eligibility rules means that

even the cream consists of relatively disadvantaged individuals. Thus, at least in programs like JTPA or WIA, cream skimming does not generally mean that program resources were spent on, for example, middle-class people. Second, in the United States, many federal, state, local, and voluntary sector employment and training programs coexist at any given point in time. For example, National Commission for Employment Policy (1995) documents that dozens of other programs coexisted with JTPA. When thinking about cream skimming in a particular program, such as WIA, it should be kept in mind that other programs may provide services better suited to the hardest to serve among that program's eligible population. Determining whether cream skimming, should it occur, is good or bad, requires more thought than the literature typically devotes to it.

13. The finding in Heckman, Smith, and Clements (1997) will hold in general for programs with small mean impacts relative to the mean untreated outcome, so long as the impacts are not strongly negatively correlated with the untreated outcome.
14. In models with regressors, this assumption is $\Delta_{a,i}(X) = Y^1_{a,i} - Y^0_{a,i} = \Delta_a$ for all i, yielding equal impacts for all persons with the same X.
15. See JTPA Directive No. D98-15, February 22, 1999.
16. In this heuristic problem, we assume that the second order conditions are satisfied.

References

Anderson, Kathryn, Richard Burkhauser, Jennie Raymond, and Clifford Russell. 1992. "Mixed Signals in the Job Training Partnership Act." *Growth and Change* 22(3): 32–48.

Barnow, Burt. 1992. "The Effects of Performance Standards on State and Local Programs." In *Evaluating Welfare and Training Programs*, Charles Manski and Irwin Garfinkel, eds. Cambridge, MA: Harvard University Press, pp. 277–309.

Barnow, Burt S., and Daniel B. Gubits. 2002. "Review of Recent Pilot, Demonstration, Research, and Evaluation Initatives to Assist in the Implementation of Programs under the Workforce Investment Act." ETA Occasional Paper No. 2003-10. Washington, DC: USDOL.

Barnow, Burt, and Jeffrey Smith. 2004. "Performance Management of U.S. Job Training Programs: Lessons from the Job Training Partnership Act." *Public Finance and Management* 4(3): 247–287.

Bell, Stephen, and Larry Orr. 2002. "Screening (and Creaming?) Applicants to Job Training Programs: The AFDC Homemaker–Home Health Aide Demonstration." *Labour Economics* 9(2): 279–301.

Bloom, Howard, Larry Orr, Stephen Bell, George Cave, Fred Doolittle, Winston Lin, and Johannes Bos. 1997. "The Benefits and Costs of JTPA Title II-A Programs: Key Findings from the National Job Training Partnership Act Study." *Journal of Human Resources* 32(3): 549–576.

Bloom, Howard, Larry Orr, George Cave, Stephen Bell, and Fred Doolittle. 1993. *The National JTPA Study: Title IIA Impacts on Earnings and Employment at 18 Months*. Research and Evaluation Report Series 93-C. Bethesda, MD: Abt Associates.

Carneiro, Pedro, Karsten Hansen, and James Heckman. 2003. "Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College." *International Economic Review* 44(2): 361–442.

Courty, Pascal, Do Han Kim, and Gerald Marschke. Forthcoming. "Curbing Cream Skimming: Evidence on Enrollment Incentives." *Labour Economics*.

D'Amico, Ron, and Jeffrey Salzman. 2004. *An Evaluation of the Individual Training Account/Eligible Training Provider Demonstration*. Social Policy Research Associates final interim report to the USDOL-ETA. Oakland, CA: Social Policy Research Associates.

Dixit, Avinash. 2002. "Incentives and Organizations in the Public Sector: An Interpretative Review." *Journal of Human Resources* 37(4): 696–727.

Djebbari, Habiba, and Jeffrey Smith. 2008. "Heterogeneous Program Impacts: Experimental Evidence from the PROGRESA Program." *Journal of Econometrics* 145(1–2): 64–80.

Doolittle, Fred, and Linda Traeger. 1990. *Implementing the National JTPA Study*. New York: Manpower Demonstration Research Corporation.

Dyke, Andrew, Carolyn Heinrich, Peter Mueser, and Kenneth Troske. 2006. "The Effects of Welfare-to-Work Program Activities on Labor Market Outcomes." *Journal of Labor Economics* 24(3): 567–608.

Fredrikkson, Peter, and Per Johansson. 2008. "Dynamic Treatment Assignment: The Consequences for Evaluations Using Observational Data." *Journal of Business and Economic Statistics* 26(4): 435–445.

General Accounting Office (GAO). 2002. *Improvements Needed in Performance Measures to Provide a More Accurate Picture of WIA's Effectiveness*. Report to Congressional Requesters. GAO-02-275. Washington, DC: GAO.

Hanushek, Eric. 2002. "Publicly Provided Education." In *Handbook of Public Finance*, Alan Auerbach and Martin Feldstein, eds. Amsterdam: North-Holland, pp. 2045–2141.

Hayek, Frederick. 1945. "The Use of Knowledge in Society." *American Economic Review* 35(4): 519–530.

Heckman, James. 1992. "Randomization and Social Program Evaluation." In *Evaluating Welfare and Training Programs*, Charles Manski and Irwin Garfinkel, eds. Cambridge, MA: Harvard University Press, pp. 201–230.

Heckman, James, Carolyn Heinrich, and Jeffrey Smith. 2002. "The Performance of Performance Standards." *Journal of Human Resources* 37(4): 778–811.

Heckman, James, Robert LaLonde, and Jeffrey Smith. 1999. "The Economics and Econometrics of Active Labor Market Programs." In *Handbook of Labor Economics,* Vol. 3A, Orley Ashenfelter and David Card, eds. Amsterdam: North-Holland, pp. 1865–2097.

Heckman, James, Lance Lochner, and Christopher Taber. 1998. "Explaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Labor Earnings with Heterogeneous Agents." *Review of Economic Dynamics* 1(1): 1–58.

Heckman, James, and Salvador Navarro. 2007. "Dynamic Discrete Choice and Dynamic Treatment Effects." *Journal of Econometrics* 136(2): 341–396.

Heckman, James, and Jeffrey Smith. 1999. "The Pre-Programme Dip and the Determinants of Participation in a Social Programme: Implications for Simple Programme Evaluation Strategies." *Economic Journal* 109(457): 313–348.

Heckman, James, Jeffrey Smith, and Nancy Clements. 1997. "Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts." *Review of Economic Studies* 64(4): 487–535.

Heckman, James, Jeffrey Smith, and Christopher Taber. 1996. "What Do Bureaucrats Do? The Effects of Performance Standards and Bureaucratic Preferences on Acceptance into the JTPA Program." In *Advances in the Study of Entrepreneurship, Innovation, and Economic Growth,* Vol. 7: *Reinventing Government and the Problem of Bureaucracy*, Gary Libecap, ed. Greenwich, CT: JAI Press, pp. 191–218.

Heinrich, Carolyn. 1999. "Do Governments Make Effective Use of Performance Management Information?" *Journal of Public Administration Research and Theory* 9(3): 363–393.

———. 2007. "False or Fitting Recognition? The Use of High Performance Bonuses in Motivating Organizational Achievements." *Journal of Policy Analysis and Management* 26(2): 281–304.

Hotz, V. Joseph, Guido Imbens, and Jacob Klerman. 2006. "Evaluating the Effects of Alternative Welfare-to-Work Training Components: A Reanalysis of the California GAIN Program." *Journal of Labor Economics* 24(3): 521–566.

Lazear, Edward. 1995. *Personnel Economics*. Cambridge, MA: MIT Press.

Lechner, Michael, Ruth Miquel, and Conny Wunsch. 2004. "Long-Run Effects of Public Sector Sponsored Training in West Germany." IZA Discussion Paper No. 1443. Bonn, Germany: IZA.

Lechner, Michael, and Jeffrey Smith. 2007. "What is the Value Added by Caseworkers?" *Labour Economics* 14(2): 135–151.

Lise, Jeremy, Shannon Seitz, and Jeffrey Smith. 2004. "Equilibrium Policy Experiments and the Evaluation of Social Programs." NBER Working Paper No. 10283. Cambridge, MA: National Bureau of Economic Research.

National Commission for Employment Policy. 1995. *Understanding Federal Training and Employment Programs*. Washington, DC: National Commission for Employment Policy.

Orr, Larry, Howard Bloom, Stephen Bell, Fred Doolittle, Winston Lin, and George Cave. 1996. *Does Training for the Disadvantaged Work? Evidence from the National JTPA Study*. Washington, DC: Urban Institute Press.

Prendergast, Canice. 1999. "The Provision of Incentives in Firms." *Journal of Economic Literature* 37(1): 7–63.

Radin, Beryl. 2006. *Challenging the Performance Movement: Accountability, Complexity, and Democratic Values*. Washington, DC: Georgetown University Press.

Sianesi, Barbara. 2004. "An Evaluation of the Swedish System of Active Labor Market Programs in the 1990s." *Review of Economics and Statistics* 86(1): 133–155.

Smith, Jeffrey. 2004. "Evaluating Local Economic Development Policies: Theory and Practice." In *Evaluating Local Economic and Employment Development: How to Assess What Works among Programmes and Policies*, Alistair Nolan and Ging Wong, eds. Paris: Organisation for Economic Co-operation and Development, pp. 287–332.

Smith, Jeffrey, Alexander Whalley, and Nathaniel Wilcox. 2010a. "Evaluating Participant Self-Evaluation." Book manuscript under preparation.

———. 2010b. "Are Participants Good Evaluators?" Unpublished manuscript. University of Michigan, Ann Arbor, MI.

Social Policy Research Associates. 2004. *The Workforce Investment Act after Five Years: Results from the National Evaluation of the Implementation of WIA*. Report prepared for the USDOL-ETA, Office of Policy Development, Evaluation and Research. Oakland, CA: SPRA.

Sosin, Michael. 1986. "Legal Rights and Welfare Change, 1960–1980." In *Fighting Poverty: What Works and What Doesn't?* Sheldon Danziger and Daniel Weinberg, eds. Cambridge, MA: Harvard University Press, pp. 260–282.

Trutko, John, Burt Barnow, Mary Farrell, and Asaph Glosser. 2005. *Earnings Replacement Outcomes for Dislocated Workers: Extent of Variation* *and Factors Accounting for Variation in Earnings Replacement Outcomes across State and Local Workforce Investment Boards*. Final report to the U.S. Department of Labor. Washington, DC: USDOL, Employment and Training Administration.

Wilson, James. 1989. *Bureaucracy: What Government Agencies Do and Why They Do It*. New York: Basic Books.